DOCUMENT RESUME

ED 315 446 TM 014 472

TITLE Maryland Functional Reading Test-II: Fall 1988 Administration. Report R-88. Technical Report.
INSTITUTION Maryland State Dept. of Education, Baltimore. Div. of Instruction.
PUB DATE Jun 89
NOTE 24p.
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Grade 9; *Graduation Requirements; High Schools; *High School Students; Public Schools; Reading Achievement; Reading Skills; *Reading Tests; *State Programs; Testing Programs; *Test Results
IDENTIFIERS *Maryland Functional Reading Test

ABSTRACT

The Maryland Functional Reading Test, Level II (MFRT-II) assesses Maryland students' mastery of functional reading skills that are described in the Declared Competencies Index published by the Maryland State Department of Education (MSDE) in 1980. The test is also used to certify students for graduation from high school and to identify those who need assistance with reading skills. Forms 170 through 179 of the 1988 edition of the MFRT-II were administered during the period October 17 through November 4, 1988 to 46,108 students in grade 9 taking the test for the first time, as well as to some students in grades 10 through 12 who were transfers or had failed the test previously. Separate sections of this paper describe and report on the development procedures, technical characteristics, and student performance results. Additional technical information and information on test reliability and validity is available from the MSDE. The percentage of students who passed on the first attempt stayed essentially the same as for the 1986 and 1987 tests (94% as opposed to 92.2% and 93.1% for 1986 and 1987, respectively). The mean scale score of 374.98 decreased slightly from the 1987 mean scale score of 376.10. Ten tables present test results and passing percentages by grade. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED315446

TM

TECHNICAL REPORT

MARYLAND FUNCTIONAL READING TEST-II
FALL 1988 ADMINISTRATION
REPORT R-88
JUNE 1989

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. A. Clenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch
200 W. Baltimore Street
Baltimore, Maryland 21201
(301) 333-2369

MFTP

MARYLAND FUNCTIONAL TESTING PROGRAM

A COMPONENT OF PROJECT BASIC FOR:

* ASSESSMENT OF BASIC AND CITIZENSHIP SKILLS
* IDENTIFICATION FOR APPROPRIATE ASSISTANCE
* CERTIFICATION FOR GRADUATION

TM014472

BEST COPY AVAILABLE

STATE BOARD OF EDUCATION

Lawrence A. Shulman, President	Rockville	1989
Wilson H. Parran, Vice President	Huntingtown	1991
May B. Bolt	Frostburg	1989
Herbert Fincher	Salisbury	1993
Priscilla Hilliard	Upper Marlboro	1990
Rosetta G. Kerr	Baltimore	1991
Joan Clark Maynard	Linthicum Heights	1992
Frederick K. Schoenbrodt	Ellicott City	1990
Benjamin Swinson	Silver Spring	1992
Shirley Leung, Student Member	Bel Air	1989

MARYLAND STATE DEPARTMENT OF EDUCATION

Joseph L. Shilling
State Superintendent of Schools

Claud E. Kitchens
Deputy State Superintendent of Schools

Martha J. Fields
Assistant Deputy State Superintendent

Nicholas Hobar
Assistant State Superintendent, Division of Instruction

James W. Latham
Director, Division of Instruction

Robert E. Gabrys
Chief, Program Assessment, Evaluation, and Instructional Support Branch

Steven Ferrara
Chief, Measurement, Statistics, and Evaluation Section

Leslie A. Walker-Bartnick
Specialist in Research and Evaluation

STATE OF MARYLAND

William Donald Schaefer
Governor

The Maryland State Department of Education does not discriminate on the basis of race, color, sex, age, national origin, religion, or handicapping conditions in matters affecting employment or in providing access to programs. For inquiries related to Departmental policy, contact the Office of Equal Opportunity.

MSDE
June 1989

Technical Report

MARYLAND FUNCTIONAL READING TEST-II
Fall 1988 Administration

Maryland State Department of Education
Division of Instruction
Program Assessment, Evaluation, and Instructional Support Branch

June 1989

The purposes of the Maryland Functional Reading Test, Level II (MFRT-II) are threefold. First, the test has been designed to assess Maryland students' mastery of functional reading skills that are described in the Declared Competencies Index (DCI) published by the Maryland State Department of Education (MSDE) in 1980. In addition, the test is used to certify students for high school graduation. Finally, the test is used to identify students who should participate in appropriate assistance classes in order to improve their reading skills.

Forms 170-179 of the 1988 edition of the Maryland Functional Reading Test, Level II were administered in Maryland public schools to students in grades 9-12 during the period October 17-November 4, 1988. Students in all of the state's 24 school districts took the MFRT-II for the first time in ninth grade. Students in grades 10, 11, and 12 who took the 1988 MFRT-II include those who failed the test in previous years as well as transfer students. The graduating class of 1982 was the first class required to pass the MFRT-II in order to receive a Maryland High School Diploma.

The remainder of this report contains separate sections describing and reporting the development procedures, technical characteristics, and student performance results of the 1988 edition of the MFRT-II. More detailed descriptions of the procedures used in test development, as well as more technical descriptions of psychometric methods used to evaluate functional tests, are available from MSDE.

DEVELOPMENT

Declared Competencies Index and the MFRT-II

All Maryland public schools are required to teach the functional skills contained in the state board endorsed Declared Competencies Index (DCI). The DCI provides the basis for the test specifications as well as for the objectives assessed by the MFRT-II. In June 1982, staff from the Division of Instruction (DOI) and Project Basic produced the Project Basic Instructional Guide, Volume VI: Functional Reading (MSDE, 1982). The document

was designed to help teachers and supervisors develop instructional activities to prepare students for the MFRT-II and describes the test items found in each of the test's five domains.

Test Specifications

The Test Specifications (MSDE, 1982) for the MFRT-II define the eligible content for the five domains of the MFRT-II that include: following directions, locating information, main idea, using details, and understanding forms. The test specifications describe the type of items used in each domain and provide a guide to item writers in generating new items for field testing and use on subsequent forms of the MFRT-II. This document is periodically reviewed and revised according to the general procedures that are available from MSDE. The test specifications committee for the MFRT-II did not meet prior to the development of the 1988 test, as no revisions were proposed or made to the test specifications during that time.

Item Development

The item writing committee for the MFRT-II is composed of educators from Maryland's public schools appointed by the State Superintendent of Schools. When necessary, the item writing committee convenes to generate new items to assess the objectives in each of the five domains in the test. Items are carefully checked for readability, item/objective match, and conformity with test specifications. They are reviewed, revised, and field tested prior to inclusion in the MFRT-II item bank. Although no new items were generated by the committee for the 1988 MFRT-II, several existing items from the item bank were revised, reviewed, and field tested again on the 1988 MFRT-II.

Item Review

In order to be included on an edition of a Maryland functional test, items must pass rigorous content, bias, psychometric, and editorial reviews. For further details concerning these procedures, contact MSDE. The operational and field test items included in the 1988 MFRT-II were submitted to these reviews.

Test Forms

Each year, ten forms of the MFRT-II are administered. All ten forms have the same 60 operational items that count toward a student's total score on the test. Each form of the test also includes ten field test items. This format allows the Program Assessment, Evaluation, and Instructional Support Branch (PAEISB) to gather field test data on a total of 100 items each year. For more detailed information on field test procedures and the purposes of field testing, contact MSDE.

TECHNICAL CHARACTERISTICS

Reliability

The reliability of the MFRT-II is evaluated using two statistical indicators: classification consistency and internal consistency coefficients. The MFRT-II is accompanied by a passing score adopted by the Maryland State Board of Education in the early 1970s. The standard for current editions of the MFRT-II (i.e., forms 110 and higher) was established by equating the MFRT-II to the original standard on Form-B of the Maryland Functional Reading Test. This standard corresponds to answering approximately 62 percent of the operational items correctly on a given test and signifies the level of performance required for graduation certification. The standard setting procedures for the MFRT-II are discussed in the MSDE publication entitled Setting a Standard for Standard Setting, (Chafin, 1983). Students whose MFRT-II score is above the passing score are considered to have mastered this set of functional skills, and students scoring below the passing score are considered not to have mastered the content and, therefore, in need of remedial instruction. The reliability or consistency with which the MFRT-II classifies students into these two categories is essential for drawing valid, believable conclusions about a student's mastery of functional reading. An indicator of the consistency of this classification is an adaptation of a procedure developed by Subkoviak (in Berk, 1984, pp.278-280), and referred to here as the PG coefficient.

A necessary condition for high classification consistency is the internal consistency of the test. Scores on criterion-referenced tests (CRTs) like the MFRT-II are used to infer the degree to which students have mastered a domain of knowledge and skill. To make valid inferences from the MFRT-II to the domain of reading knowledge and skill defined by the DCI, items must measure only this knowledge and skill. Internal consistency reliability coefficients provide a measure of the extent to which the items on a test are homogeneous, or measure the same thing, thereby providing one way of describing test reliability. One internal consistency reliability coefficient is the Kuder-Richardson formula number 20, or KR20 (see, for example, Crocker & Algina, 1986; Gronlund, 1982). The reliability and classification consistency of the 1988 MFRT-II are discussed in the following sections.

Classification consistency. Table 1 indicates that the classification consistency of the MFRT-II is excellent: .97 for students who took the test in 1988. This coefficient reveals that approximately 97 percent of the students who took the test would be classified in the same way (as a passer or failer) on a second concurrent administration of the test. The PG of .97 indicates that only about 3 percent of the students who took the test might have been misclassified as passers or failers. It is

likely that students who may have been classified incorrectly are those whose total scale score was just above or just below the passing score of 340.

Table 1

Classification Consistency (PG) and Internal Consistency (KR20) Coefficients

All forms, all grade 9 students (N=46,108)

Domain	No. Items	KR20	PG
1. Following Directions	13	.68	.90
2. Locating Information	13	.70	.90
3. Main Ideas	10	1.08[a]	.91
4. Using Details	13	.61	.94
5. Understanding Forms	11	.74	.92
Total Test	60	.90	.97

Individual test forms (70 items each form[b])

Form	No. Students	KR20
170	4,567	.80
171	4,647	.82
172	4,583	.83
173	4,280	.82
174	4,595	.83
175	4,545	.83
176	4,666	.83
177	4,618	.82
178	4,515	.81
179	4,580	.82

Note. [a]The normal range of reliability coefficients is from 0 to 1. The KR20 of 1.08 for domain 3 is a computational artifact which can occur because of the limited variance of scores in domain 3. The KR20 for domain 3 indicates that internal consistency is very high for this domain. [b]KR20 coefficients are not available at this time for the 60 operational items.

Internal consistency. Table 1 contains the internal consistency coefficients of the 1988 MFRT-II forms and items across forms. KR20 coefficients for the individual forms have remained at a stable and high rate over the last four years. On the 1988 test, they ranged from .80 to .83. Although the number of items in each domain is limited (ranging from 10 to 13), KR20 coefficients are satisfactory for all domains, ranging from .61 to 1.08 on the 1988 test. These coefficients have remained stable over the last four years. The KR20 coefficient for the total test is higher than those for individual domains because domain coefficients are calculated on fewer items than for the total test. KR20s are lower bound estimates of test reliability.

Validity

The primary aim in developing, administering, scoring, and reporting the MFRT-II is to provide accurate decisions about students' possession of reading skills. The first condition necessary to assure the validity of any test is the demonstration of acceptable levels of reliability and classification consistency as described in the previous section. Evidence of the validity of the MFRT-II includes information about content validity, construct validity, and freedom from bias. Complete descriptions of the procedures used to help assure that the MFRT-II is administered according to prescribed procedures, methods used to determine content and construct validity, and procedures used to assist in determining that items are free of bias are available from MSDE.

Equivalence of Forms

The issue of equity, or fairness, is of major importance to any test that has a designated passing score and is used to make decisions concerning high school graduation. Students who take the MFRT-II must be assured that the test they are taking is no more easy or difficult to pass than other editions or forms of the test. Thus, equivalence of test forms refers to equivalence of the difficulty of passing a particular form of the test.

Items are selected for inclusion on a given edition of the MFRT-II to help ensure the equivalence of current forms to previous editions. Statistical pre-equating analysis is performed before the test is administered to ensure that the forms are equivalent in difficulty to one another and to previous years' editions. This analysis also provides a way to adjust the passing score in order to assure comparability of the standard. The details of PAEISB's Rasch equating procedure are available from MSDE.

Table 2 shows raw scores (number of items correct) and their corresponding equated scale, domain, and percent correct scores and standard errors. Students who correctly answered 38 items on the 1988 MFRT-II achieved a scale score of 341, which is the

Table 2

Lookup Table for the Total Test

	Raw Score		Domain Score		Percent Score		Scale Score		Scale Score Standard Error	
	0		0		0		217		30	
	1		1		1		237		20	
	2		2		3		251		15	
	3		4		5		260		12	
	4		5		6		267		11	
	5		7		8		272		10	
	6		8		10		278		9	
	7		10		11		280		8	
	8		11		13		284		8	
	9		13		15		287		8	
	10		14		16		289		7	
	11		16		18		292		7	
	12		17		20		295		7	
	13		19		21		297		7	
	14		20		23		299		7	
	15		21		25		301		6	
	16		23		26		303		6	
	17		24		28		305		6	
	18		26		30		307		6	
	19		27		31		309		6	
	20		29		33		311		6	
	21		31		35		313		6	
	22		31		36		314		6	
	23		33		38		316		6	
	24		35		40		318		6	
	25		36		41		319		6	
	26		38		43		321		6	
	27		39		45		323		6	
	28		40		46		324		6	
	29		42		48		326		6	
	30		44		50		328		6	
	31		45		51		329		6	
	32		47		53		331		6	
	33		48		54		332		6	
	34		50		56		334		6	
	35		52		58		336		6	
	36		53		60		337		6	
	37		55		61		339		6	
Pass Score *	38	**	56	**	63	**	341	*	6	* Pass Score
	39		57		65		342		6	
	40		59		66		344		6	
	41		61		68		346		6	
	42		63		70		348		6	
	43		64		71		350		6	
	44		66		73		352		6	
	45		68		75		354		6	
	46		70		76		356		7	
	47		71		78		358		7	
	48		73		80		360		7	
	49		75		81		363		7	
	50		77		83		366		7	
	51		78		85		368		8	
	52		81		86		372		8	
	53		82		88		375		8	
	54		85		90		379		9	
	55		87		91		383		10	
	56		88		93		388		11	
	57		91		95		395		12	
	58		93		96		404		15	
	59		96		98		419		20	
	60		98		100		438		30	

lowest possible passing scale score in 1988. This passing raw score is the same as the lowest possible passing raw score of 38 in 1987.

Table 3 contains raw scores, equated scale scores, domain scores, and percent correct scores and standard errors for each domain. To illustrate the use of this table, the panel for domain 1 indicates that a student who correctly answered 10 of 13 items, equivalent to a scale score of 346, can be considered to have mastered the domain. However, pass-fail status is not determined by performance within an individual domain: a student may perform poorly on a given domain yet pass the test by performing well in other domains. One reason for determining pass-fail status only at the total test level is that longer tests are generally more reliable than shorter tests. This relationship between test length and test reliability is reflected in the scale score standard errors (SEss) for domain one. A scale score of 346, which is based on 13 items in the domain, has a standard error of 14 -- higher than the standard error of 6 for a scale score of 346 based on 60 items (see Tables 2 and 3).

Item "Bias" Analysis

For the MFTP, test item "bias" is defined as differential item functioning (DIF) in contrasting groups of examinees. Potential DIF for reference and focal groups -- males versus females and whites versus blacks -- is sought in two different stages. First, during the item development stage, items are subjected to judgmental reviews by a bias committee. Any item judged to potentially function differently for one group than for another is deleted from the item pool or revised before field testing. Second, all operational and field test items are statistically analyzed for potential DIF after each census test administration.

The statistical approach used to detect DIF is the transformed item difficulties (TID) method described by Osterlind (1985, p. 34). The purpose of the analysis is to identify items that may be functioning differently for reference versus focal groups. If an item is found to be more difficult for one group than for another, the item is flagged as possibly "biased." Flagged items are put on probationary status and their content and format are reviewed at the next meeting of the bias committee for possible causes of DIF. Flagged items can be used on subsequent editions of the MFRT-II, but if they are flagged statistically on a subsequent edition as possibly "biased" they are removed from the item bank. If a flagged item is not flagged in its next two administrations, it is removed from probation.

The TID approach studies the interaction of two or more groups' performance with items. TID uses a relative standard: that is, an item is considered to be biased when it is comparatively more difficult for one group to answer correctly than it is for the other. Item p-values are calculated for each of two

Table 3

Equated Raw, Domain, and Scale Scores and Standard Errors for Each Domain

DOMAIN 1 - FOLLOWING DIRECTIONS

NUMBER OF ITEMS = 13 MEAN DIFFICULTY = .597 STD. DEVIATION = .994

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	1	0	-3.416	1.609	237	32	2.84
1	5	7	-2.265	1.073	260	21	5.20
2	11	15	-1.410	.810	277	16	7.36
3	17	23	-.835	.711	288	14	9.04
4	24	30	-.370	.653	298	13	10.24
5	30	38	.041	.627	306	13	11.23
6	38	46	.425	.614	314	12	11.88
7	45	53	.800	.612	321	12	12.22
8	53	61	1.180	.623	329	12	12.27
9	60	69	1.582	.646	337	13	11.99
10	69	76	2.034	.700	348	14	11.50
11	77	84	2.591	.798	357	16	10.58
12	85	92	3.424	1.061	373	21	9.22
13	93	100	4.550	1.592	396	32	7.00

DOMAIN 2 - LOCATING INFORMATION

NUMBER OF ITEMS = 13 MEAN DIFFICULTY = .535 STD. DEVIATION = .711

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	3	0	-3.256	1.581	240	32	5.15
1	8	7	-2.145	1.054	262	21	8.37
2	17	15	-1.328	.786	278	16	10.23
3	25	23	-.790	.685	289	14	11.13
4	33	30	-.360	.626	298	13	11.39
5	40	38	.017	.600	305	12	11.53
6	46	46	.389	.587	312	12	11.45
7	53	53	.712	.587	319	12	11.27
8	60	61	1.062	.599	328	12	11.00
9	66	69	1.436	.624	334	12	10.61
10	73	76	1.862	.682	342	14	10.22
11	81	84	2.394	.782	353	16	9.50
12	88	92	3.204	1.050	369	21	8.28
13	95	100	4.307	1.575	391	32	5.71

DOMAIN 3 - MAIN IDEA

NUMBER OF ITEMS = 10 MEAN DIFFICULTY = 2.397 STD. DEVIATION = .573

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	2	0	-1.054	1.600	284	32	5.31
1	9	10	.083	1.066	307	21	9.07
2	19	20	.924	.807	323	16	11.78
3	29	30	1.493	.705	335	14	13.24
4	38	40	1.963	.667	344	13	14.21
5	47	50	2.398	.655	353	13	14.57
6	57	60	2.832	.667	362	13	14.45
7	67	70	3.302	.705	371	14	13.76
8	78	80	3.870	.806	382	16	12.63
9	88	90	4.708	1.066	399	21	10.33
10	95	100	5.844	1.598	422	32	6.60

Table 3 (Cont.)

Equated Raw, Domain, and Scale Scores and Standard Errors for Each Domain

DOMAIN 4 - USING DETAILS

NUMBER OF ITEMS = 13 MEAN DIFFICULTY = 1.068 STD. DEVIATION = .902

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	1	0	-2.788	1.579	250	32	3.08
1	5	7	-1.658	1.053	272	21	5.47
2	11	15	-.843	.785	288	16	7.49
3	18	23	-.309	.683	299	14	8.96
4	23	30	.118	.626	307	13	10.01
5	30	38	.494	.600	315	12	10.93
6	36	46	.846	.590	322	12	11.63
7	44	53	1.195	.593	329	12	12.20
8	51	61	1.556	.611	336	12	12.59
9	59	69	1.952	.649	344	13	12.82
10	67	76	2.414	.715	353	14	12.68
11	77	84	3.007	.827	365	17	11.70
12	87	92	3.920	1.107	383	22	9.51
13	94	100	5.147	1.661	408	33	6.22

DOMAIN 5 - UNDERSTANDING FORMS

NUMBER OF ITEMS = 11 MEAN DIFFICULTY = 1.369 STD. DEVIATION = .810

RAW SCORE	DOMAIN SCORE	PERCENT OF TEST	LOGIT ABILITY	LOGIT STD.ERROR	SCALED ABILITY	SCALED STD.ERROR	DOMAIN SCORE STD. ERROR
0	3	0	-2.233	1.607	260	32	5.71
1	10	9	-1.086	1.071	283	21	9.22
2	20	18	-.238	.806	300	16	11.18
3	29	27	.326	.704	312	14	12.03
4	38	36	.782	.650	321	13	12.25
5	44	45	1.189	.628	329	13	12.23
6	53	54	1.580	.626	337	13	12.02
7	60	63	1.982	.644	345	13	11.67
8	67	72	2.425	.693	354	14	11.24
9	75	81	2.970	.792	364	16	10.60
10	84	90	3.788	1.056	381	21	9.65
11	92	100	4.903	1.584	403	32	7.71

groups that have been matched on ability. These p-values are converted to z-values and then transformed to delta values with a mean of 13 and a standard deviation of 4. The delta values for each pair of comparison groups for each item are plotted. Potential DIF is sought by (1) examining 99 percent confidence intervals constructed around a regression line for each plot, (2) examining standardized DIF values, and (3) inspecting plots visually. "DIF" is defined as the distance of each plotted item from the major axis line. Unacceptable "DIF" is defined in the MFTP as any item with a DIF value outside of the 99 percent confidence interval, or a standardized DIF value greater than plus or minus two (Angoff, 1982, p. 98; Rudner, 1977). Plots provide descriptive information that aids in interpretation of DIF results.

DIF for males versus females. Table 4 and Figure 1 contain the 1988 MFRT-II DIF information for males (reference group) versus females (focal group). DIF and standardized DIF values can be found in the two right hand columns in Table 4 under the heading "DIF*" and "STD. DIF." Any item with a DIF value outside of the 99 percent confidence interval or a standardized DIF value greater than plus or minus two may be functioning differently for one group than it is for another and is flagged for further study. DIF values between -.84 and .84 are within 99 percent confidence limits for the male-female comparisons. Non-significant differences between values are likely due to random fluctuations in subgroup performance and measurement error.

Only one DIF value in Table 4 falls outside of the 99 percent confidence interval; this item also has a standardized DIF value greater than 2.0. Operational sequence number 52, (actual item number 60) has a DIF value of .92 and a standardized DIF of -2.83. In addition, operational sequence numbers 2, 10, 13, and 36 (actual item numbers 2, 11, 14, and 41 respectively) have standardized DIF values outside of the acceptable limits although the DIF values are within acceptable limits (see Table 4). Items 2, 10, 13 and 36 were slightly more difficult for females than for males, and item 52 was more difficult for males than for females. These five items have been put on probationary status and will be discussed in future reports. The DIF criterion has resulted in identification of five items with statistical evidence of possible "bias" but still fall within the item cluster on the plot. Figure 1 shows the plot of the standardized delta values; however no points on the plot are outside of the cluster. Finally, the correlation between the two groups' DIF values for all items is .98, indicating a strong degree of linear relationship.

DIF for whites versus blacks. Table 5 and Figure 2 contain the DIF information for whites (reference group) versus blacks (focal group). For this comparison, the 99 percent confidence interval limits are -.77 and .77. There were no items with DIF values outside of these acceptable limits although operational sequence numbers 6 and 10 (actual item numbers 6 and 11, respectively) do have standardized DIF values outside of the acceptable

Table 4

DIF and Standardized DIF Values for Bias Analysis: Males versus Females

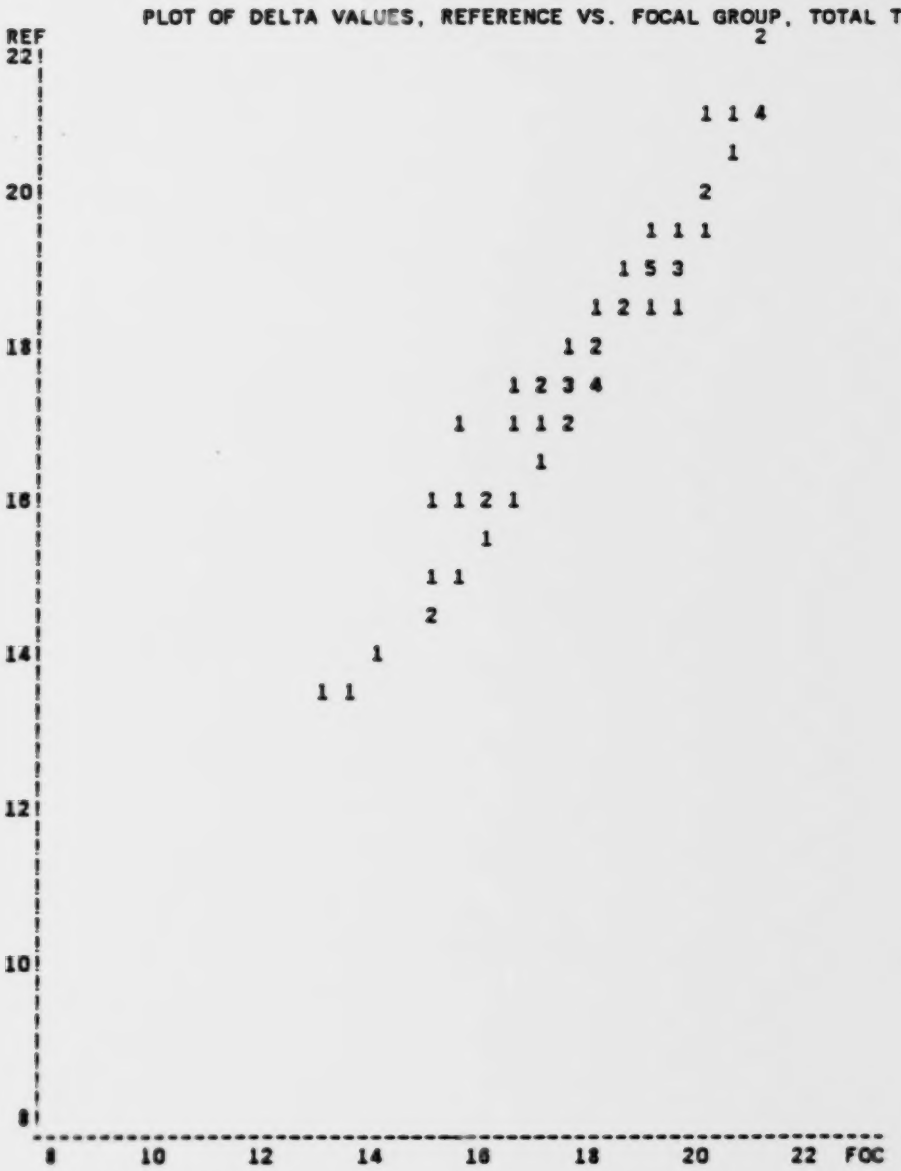
SEQUENCE NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	DIF*	STD. DIF
I001	.97	.97	1.88	1.88	20.52	20.52	.08	.26
I002	.98	.96	2.05	1.75	21.20	20.00	-.78	-2.35
I003	.86	.82	1.08	.92	17.32	16.68	-.46	-1.42
I004	.77	.71	.74	.55	15.96	15.20	-.58	-1.80
I005	.93	.93	1.48	1.48	18.92	18.92	.04	.13
I006	.79	.77	.81	.74	16.24	15.96	-.23	-.71
I007	.87	.87	1.13	1.13	17.52	17.52	.00	.01
I008	.96	.96	1.75	1.75	20.00	20.00	.07	.22
I009	.88	.88	1.23	1.17	17.92	17.68	-.16	-.49
I010	.99	.98	2.33	2.05	22.32	21.20	-.67	-2.08
I011	.92	.92	1.41	1.41	18.64	18.64	.03	.11
I012	.87	.85	1.13	1.04	17.52	17.16	-.28	-.79
I013	.99	.98	2.33	2.05	22.32	21.20	-.67	-2.08
I014	.87	.85	1.13	1.04	17.52	17.16	-.26	-.79
I015	.58	.58	.15	.15	13.80	13.80	-.10	-.31
I016	.67	.70	.44	.52	14.76	15.08	.16	.50
I017	.83	.86	.95	1.08	16.80	17.32	.36	1.11
I018	.70	.73	.52	.61	15.08	15.44	.20	.61
I019	.88	.89	1.17	1.23	17.68	17.92	.18	.56
I020	.83	.84	.95	.99	16.80	16.96	.10	.31
I021	.58	.60	.20	.25	13.80	14.00	.05	.15
I022	.66	.69	.41	.50	14.64	15.00	.19	.57
I023	.76	.75	.71	.67	15.84	15.68	-.18	-.48
I024	.85	.87	1.04	1.13	17.16	17.52	.25	.78
I025	.94	.92	1.55	1.41	19.20	18.64	-.35	-1.09
I026	.98	.98	2.05	2.05	21.20	21.20	.10	.32
I027	.88	.90	1.17	1.28	17.68	18.12	.33	1.00
I028	.93	.93	1.48	1.48	18.92	18.92	.04	.13
I029	.93	.94	1.48	1.55	18.92	19.20	.24	.75
I030	.96	.96	1.75	1.75	20.00	20.00	.07	.22
I031	.98	.98	2.05	2.05	21.20	21.20	.10	.32
I032	.87	.88	1.13	1.17	17.52	17.68	.12	.37
I033	.88	.88	1.17	1.17	17.68	17.68	.01	.03
I034	.88	.87	2.05	1.88	21.20	20.52	-.39	-1.19
I035	.92	.90	1.41	1.28	18.64	18.12	-.34	-1.05
I036	.83	.75	.95	.67	16.80	15.68	-.82	-2.53
I037	.93	.94	1.48	1.55	18.92	19.20	.24	.75
I038	.91	.93	1.34	1.48	18.36	18.92	.43	1.33
I039	.98	.98	2.05	2.05	21.20	21.20	.10	.32
I040	.95	.93	1.65	1.48	19.60	18.92	-.43	-1.32
I041	.94	.95	1.55	1.65	19.20	19.60	.34	1.04
I042	.95	.95	1.65	1.65	19.60	19.60	.06	.18
I043	.56	.52	.15	.05	13.60	13.20	-.39	-1.20
I044	.89	.90	1.23	1.28	17.92	18.12	.16	.49
I045	.82	.84	.92	.99	16.68	16.96	.18	.57
I046	.83	.82	.95	.92	16.80	16.68	-.10	-.31
I047	.94	.94	1.55	1.55	19.20	19.20	.05	.15
I048	.94	.95	1.55	1.65	19.20	19.60	.34	1.04
I049	.94	.95	1.55	1.65	19.20	19.60	.34	1.04
I050	.89	.89	1.23	1.23	17.92	17.92	.01	.05
I051	.98	.98	2.05	2.05	21.20	21.20	.10	.32
I052	.91	.95	1.34	1.65	18.36	19.60	.92	2.83
I053	.88	.89	1.17	1.23	17.68	17.92	.18	.56
I054	.95	.96	1.65	1.75	19.60	20.00	.35	1.07
I055	.79	.78	.81	.77	16.24	16.08	-.15	-.45
I056	.91	.91	1.34	1.34	18.36	18.36	.03	.08
I057	.88	.89	1.17	1.23	17.68	17.92	.18	.56
I058	.70	.69	.52	.50	15.08	15.00	-.12	-.37
I059	.79	.82	.81	.92	16.24	16.68	.29	.88
I060	.75	.79	.67	.81	15.68	16.24	.36	1.10

TOTAL RECORDS-REFERENCE = 20157
TOTAL RECORDS-----FOCAL = 20186

STANDARD DEV.-REFERENCE = 2.09
STANDARD DEV.-----FOCAL = 2.01

VARIANCE------REFERENCE = 4.36
VARIANCE----------FOCAL = 4.05

MEAN DELTA----REFERENCE = 18.04
MEAN DELTA--------FOCAL = 18.02



SLOPE = 1.04
INTERCEPT = -.67
CORRELATION COEFFICIENT= .98

MEAN DIF = -.00
STANDARD DEVIATION DIF= .32

98% C.I. = -.00 + OR - 2.58 * .32 = (-.84, .84)

*DIF = DISTANCE FROM THE REGRESSION LINE

Figure 1. Plot of standardized delta values for males versus females.

Table 5

DIF and Standardized DIF Values for Bias Analysis: Whites versus Blacks

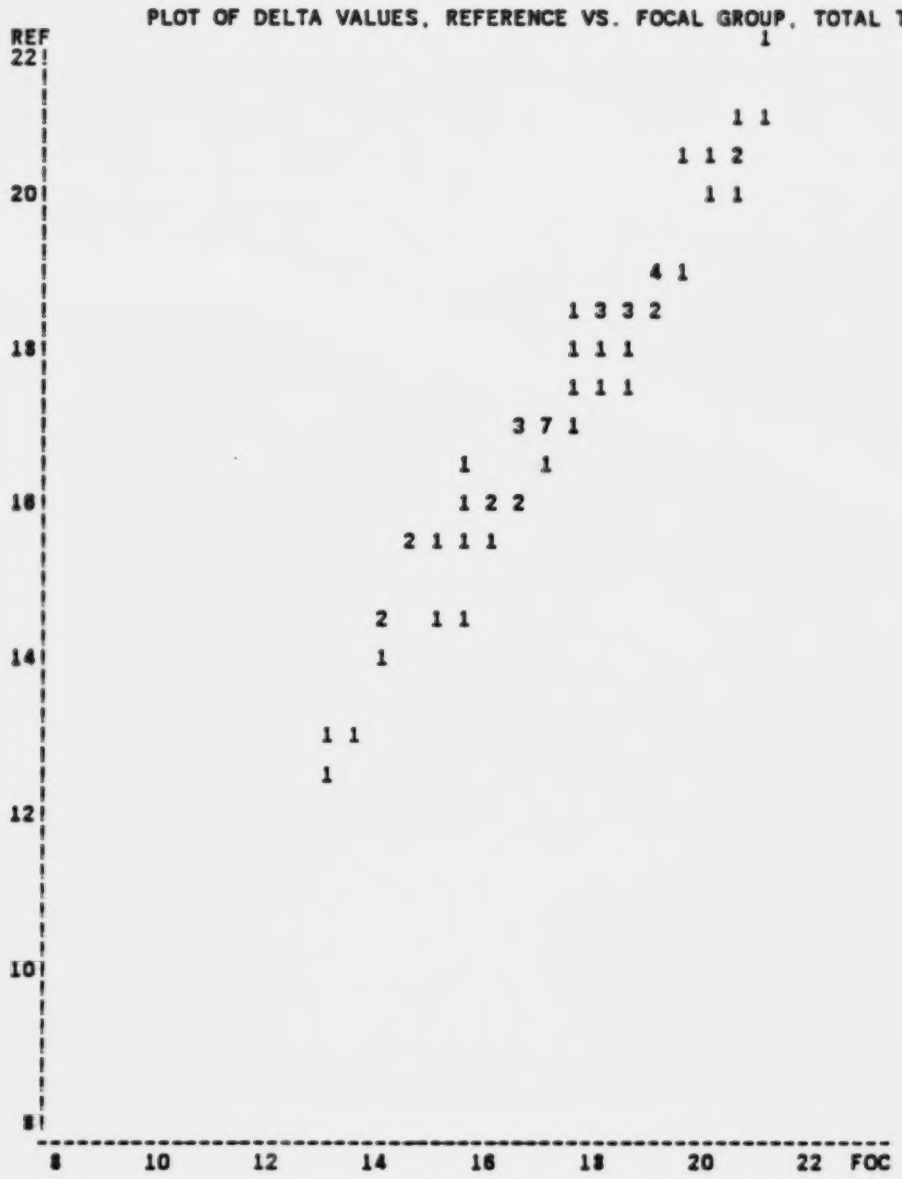
SEQUENCE NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	DIF*	STD. DIF
I001	.96	.96	1.75	1.75	20.00	20.00	.08	.27
I002	.97	.95	1.88	1.65	20.52	19.60	-.57	-1.91
I003	.81	.75	.88	.67	16.52	15.68	-.59	-1.99
I004	.72	.66	.58	.41	15.32	14.64	-.50	-1.68
I005	.91	.90	1.34	1.28	18.36	18.12	-.13	-.42
I006	.75	.67	.67	.44	15.68	14.76	-.67	-2.24
I007	.83	.82	.95	.92	16.80	16.68	-.07	-.24
I008	.94	.94	1.55	1.55	19.20	19.20	.08	.21
I009	.85	.84	1.04	.99	17.16	16.96	-.12	-.41
I010	.99	.98	2.33	2.05	22.32	21.20	-.68	-2.26
I011	.91	.88	1.34	1.17	18.36	17.68	-.44	-1.48
I012	.83	.81	.95	.88	16.80	16.52	-.19	-.62
I013	.98	.98	2.05	2.05	21.20	21.20	.10	.35
I014	.81	.83	.88	.95	16.52	16.80	.21	.70
I015	.51	.53	.03	.08	13.12	13.32	.08	.28
I016	.63	.61	.33	.28	14.32	14.12	-.18	-.60
I017	.77	.80	.74	.84	15.96	16.36	.28	.95
I018	.67	.72	.44	.58	14.76	15.32	.37	1.25
I019	.84	.83	.99	.95	16.96	16.80	-.10	-.32
I020	.77	.79	.74	.81	15.96	16.24	.20	.66
I021	.50	.52	.00	.05	13.00	13.20	.08	.27
I022	.60	.60	.25	.25	14.00	14.00	-.04	-.14
I023	.65	.68	.39	.47	14.56	14.88	.20	.66
I024	.83	.80	.95	.84	16.80	16.36	-.30	-1.01
I025	.92	.89	1.41	1.23	18.84	17.92	-.48	-1.55
I026	.98	.97	2.05	1.88	21.20	20.52	-.38	-1.28
I027	.84	.84	.99	.99	16.96	16.96	.02	.08
I028	.91	.90	1.34	1.28	18.36	18.12	-.13	-.42
I029	.91	.93	1.34	1.48	18.36	18.92	.45	1.50
I030	.94	.95	1.55	1.65	19.20	19.60	.35	1.17
I031	.97	.97	1.88	1.88	20.52	20.52	.09	.30
I032	.85	.88	1.04	1.17	17.16	17.68	.39	1.32
I033	.83	.83	.95	.95	16.80	16.80	.01	.05
I034	.97	.96	1.88	1.75	20.52	20.00	-.28	-.95
I035	.89	.87	1.23	1.13	17.92	17.52	-.25	-.83
I036	.76	.72	.71	.58	15.84	15.32	-.38	-1.27
I037	.90	.91	1.28	1.34	18.12	18.36	.21	.71
I038	.88	.91	1.17	1.34	17.68	18.36	.52	1.74
I039	.96	.97	1.75	1.88	20.00	20.52	.45	1.52
I040	.92	.92	1.41	1.41	18.84	18.84	.05	.17
I041	.92	.93	1.41	1.48	18.84	18.92	.25	.85
I042	.93	.93	1.48	1.48	18.92	18.92	.06	.19
I043	.48	.49	-.10	-.03	12.60	12.88	.13	.43
I044	.85	.85	1.04	1.04	17.16	17.16	.02	.07
I045	.77	.77	.74	.74	15.96	15.96	-.00	-.01
I046	.78	.81	.77	.88	16.08	16.52	.32	1.06
I047	.92	.91	1.41	1.34	18.84	18.36	-.15	-.50
I048	.92	.92	1.41	1.41	18.84	18.84	.05	.17
I049	.93	.94	1.48	1.55	18.92	19.20	.28	.87
I050	.86	.88	1.08	1.17	17.32	17.68	.28	.95
I051	.97	.97	1.88	1.88	20.52	20.52	.09	.30
I052	.89	.89	1.23	1.23	17.92	17.92	.04	.13
I053	.84	.83	.99	.95	16.96	16.80	-.10	-.32
I054	.93	.94	1.48	1.55	18.92	19.20	.26	.87
I055	.74	.79	.64	.81	15.56	16.24	.48	1.60
I056	.86	.89	1.08	1.23	17.32	17.92	.48	1.52
I057	.84	.85	.99	1.04	16.96	17.16	.16	.54
I058	.64	.61	.36	.28	14.44	14.12	-.26	-.88
I059	.75	.75	.67	.67	15.68	15.68	-.01	-.03
I060	.72	.71	.58	.55	15.32	15.20	-.10	-.34

TOTAL RECORDS-REFERENCE = 9823
TOTAL RECORDS-----FOCAL = 9837

STANDARD DEV.-REFERENCE = 2.12
STANDARD DEV.-----FOCAL = 2.08

VARIANCE------REFERENCE = 4.48
VARIANCE----------FOCAL = 4.24

MEAN DELTA----REFERENCE = 17.38
MEAN DELTA--------FOCAL = 17.34



SLOPE = 1.03
INTERCEPT = -.47
CORRELATION COEFFICIENT= .98

MEAN DIF = -.00
STANDARD DEVIATION DIF= .30

99% C.I. = -.00 + OR - 2.58 * .30 = (-.77, .77)

*DIF = DISTANCE FROM THE REGRESSION LINE

Figure 2. Plot of standardized delta values for whites versus blacks.

range (-2.24 and -2.26, respectively). Both items appear to be slightly easier for whites. These items have been put on probation and will be discussed in future reports. All of the remaining DIF and standardized DIF values in Table 5 fall within acceptable ranges, indicating that no additional item on the 1988 MFRT-II shows statistical evidence of potential race "bias." Figure 2 shows that all points on the plot are within the major cluster. Finally, the correlation between the two groups' DIF values for all items is .98, indicating a strong degree of linear relationship.

Analysis of Item 57

During Fall 1988 testing, it was discovered that one form of the MFRT-II had an item number missing. This occurred with item number 57, which was a field test item on Form 173. An analysis was conducted to make sure this printing error did not have an impact on student scores. The data in Table 6 are taken from the item history, and show this item's performance when it was field-tested in 1985 as well as the current year.

Table 6

Item History for Item with a Missing Item Number

Year	Total Fit	Point Biserial	N	Response Proportions A	B	C	D	Other	p
1985	-.776	.42	5035	.38	.55	.04	.02	.01	.55
1988	-.105	.38	4280	.33	.61	.03	.02	.00	.61

Regardless of the fact that the item number was missing for this item, students performed at a level consistent with previous performances, and in fact, show slight improvement. This is probably attributable to the fact that over the three years since the item was first used, instruction has improved slightly. Since the students performed well on this item, it can be assumed that the lack of an item number did not inhibit their ability to enter their answer on their answer sheet. Since the item was a field-test item, it was not counted in any student's total score.

RESULTS

Pass-rates. The percentage of students who took the MFRT-II for the first time and passed stayed essentially the same as for the 1986 and 1987 tests (see Table 7). The pass-rates for the past seven years for first time takers as well as retested students are also presented.

Table 7

Pass-Rates for the MFRT-II: 1982-1988

	Percent Passing						
	School Year						
Grade	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88	1988-89
9	89.1	93.1	93.1	92.2	92.2	93.1	94.0
10	67.1	NA	78.6	76.8	74.5	78.0	80.5
11	74.7	NA	84.2	81.3	83.0	84.6	87.2
12	78.0	NA	87.7	84.0	85.1	88.1	89.9

Note. Percentages include fall administration only.

NA Not available.

Response Proportions. Table 8 contains the response proportions and p-values for each operational item on the MFRT-II. The response proportion indicates the proportion of students selecting each possible response for each item on the MFRT-II. Response proportions are used to evaluate the quality of the incorrect responses (or distractors) and can be useful in identifying confusing or miskeyed items. The p-value indicates the percentage of students who answered an item correctly. The p-values for the 1988 MFRT-II range from .54 (operational sequence number 43, actual item number 49) to .98 (operational sequence numbers 10, 13, and 26, actual item numbers 11, 14, and 30). Experience and convention has defined the acceptable range of p-values as .30 to .70 for norm-referenced tests. However, this convention applies less strictly to criterion-referenced tests that may include items outside of this range in order to ensure adequate coverage of relevant content areas. In particular, it is appropriate for basic skills tests like the MFRT-II to contain a number of "easier" items -- with p-values greater than .70 -- to ensure accurate measurement of low-achieving students and meaningful diagnostic domain scores.

Summary Results. The mean raw, domain, and scale scores and their standard deviations for the 1988 MFRT-II are presented in Table 9. The mean scale score of 374.98 decreased slightly from 1987s mean scale score of 376.10. Finally, Table 10 presents the distribution of frequencies of each score point for the 1988 MFRT-II.

Table 8

Percentages of Students Who Answered Each Item Correctly (P-Values) and Response Proportions

SEQUENCE NUMBER	RESPONSE PROPORTIONS A	B	C	D	OTHER	P-VALUE
I001	.96	.02	.01	.01	.00	.96
I002	.01	.96	.02	.00	.00	.96
I003	.01	.02	.83	.15	.00	.83
I004	.22	.73	.04	.02	.00	.73
I005	.92	.02	.01	.05	.00	.92
I006	.78	.04	.15	.04	.00	.78
I007	.08	.04	.86	.03	.00	.86
I008	.95	.03	.01	.01	.00	.95
I009	.07	.02	.04	.87	.00	.87
I010	.98	.01	.00	.01	.00	.98
I011	.01	.01	.07	.91	.00	.91
I012	.05	.85	.04	.05	.00	.85
I013	.01	.98	.01	.01	.00	.98
I014	.02	.85	.03	.11	.00	.85
I015	.28	.56	.07	.09	.00	.56
I016	.07	.22	.03	.68	.00	.68
I017	.08	.05	.83	.08	.00	.83
I018	.04	.71	.23	.02	.00	.71
I019	.04	.87	.07	.02	.00	.87
I020	.03	.82	.10	.05	.00	.82
I021	.24	.58	.08	.09	.00	.58
I022	.87	.04	.10	.09	.00	.87
I023	.74	.08	.11	.07	.00	.74
I024	.08	.02	.84	.06	.00	.84
I025	.04	.92	.03	.01	.00	.92
I026	.98	.01	.01	.00	.00	.98
I027	.06	.87	.04	.02	.00	.87
I028	.07	.92	.01	.00	.00	.92
I029	.05	.01	.01	.93	.00	.93
I030	.01	.01	.03	.95	.00	.95
I031	.97	.01	.02	.00	.00	.97
I032	.01	.05	.87	.07	.00	.87
I033	.86	.08	.03	.02	.00	.86
I034	.01	.01	.98	.01	.00	.98
I035	.05	.90	.04	.01	.00	.90
I036	.02	.19	.01	.78	.00	.78
I037	.03	.02	.92	.02	.00	.92
I038	.01	.03	.05	.91	.00	.91
I039	.01	.97	.01	.01	.00	.97
I040	.01	.05	.93	.01	.00	.93
I041	.93	.03	.03	.01	.00	.93
I042	.01	.03	.02	.94	.00	.94
I043	.18	.54	.25	.03	.00	.54
I044	.04	.06	.87	.02	.00	.87
I045	.05	.09	.05	.81	.00	.81
I046	.81	.15	.01	.02	.00	.81
I047	.03	.01	.03	.92	.00	.92
I048	.93	.04	.02	.01	.00	.93
I049	.03	.01	.94	.02	.00	.94
I050	.88	.02	.09	.01	.00	.88
I051	.01	.97	.01	.01	.00	.97
I052	.03	.04	.01	.91	.00	.91
I053	.08	.87	.02	.03	.00	.87
I054	.01	.02	.02	.94	.00	.94
I055	.02	.02	.77	.18	.00	.77
I056	.01	.90	.05	.04	.00	.90
I057	.01	.07	.87	.05	.00	.87
I058	.68	.08	.14	.09	.00	.68
I059	.10	.79	.03	.08	.00	.79
I060	.02	.19	.03	.75	.00	.75

TOTAL RECORDS = 46108

Table 9

Mean Raw, Domain, and Scale Scores and Standard Deviations (in parentheses): Grade 9 (N=46,108)

Domain	Maximum Raw Score	Raw Score	Domain Score	Scale Score
Following Directions	13	11.6 (1.8)	81.4 (14.2)	371.5 (24.1)
Locating Information	13	11.7 (1.8)	86.1 (12.4)	370.3 (22.7)
Main Ideas	10	7.3 (2.2)	69.8 (21.1)	379.2 (26.7)
Using Details	13	11.4 (1.9)	81.2 (15.89	379.2 (25.7)
Understanding Forms	11	9.3 (1.8)	78.4 (14.9)	375.4 (19.3)
All Domains	60	51.35 (7.66)	-- --	374.98 (23.99)

Table 10

Frequency Distribution of Total Scores

RAW SCORE	SCALE SCORE	PERCENT SCORE	FREQUENCY	FREQUENCY PERCENTAGE	CUMULATIVE FREQUENCY	CUMULATIVE PERCENTAGE
0	217	.0	2	.0	2	.0
1	237	1.7	0	.0	2	.0
2	251	3.3	0	.0	2	.0
3	260	5.0	1	.0	3	.0
4	267	6.7	0	.0	3	.0
5	272	8.3	3	.0	6	.0
6	278	10.0	3	.0	9	.0
7	280	11.7	2	.0	11	.0
8	284	13.3	3	.0	14	.0
9	287	15.0	3	.0	17	.0
10	289	16.7	8	.0	25	.1
11	292	18.3	13	.0	38	.1
12	295	20.0	22	.0	60	.1
13	297	21.7	23	.0	83	.2
14	299	23.3	37	.1	120	.3
15	301	25.0	27	.1	147	.3
16	303	26.7	42	.1	189	.4
17	305	28.3	57	.1	246	.5
18	307	30.0	73	.2	319	.7
19	309	31.7	75	.2	394	.9
20	311	33.3	61	.1	455	1.0
21	313	35.0	76	.2	531	1.2
22	314	36.7	82	.2	613	1.3
23	316	38.3	75	.2	688	1.5
24	318	40.0	77	.2	765	1.7
25	319	41.7	76	.2	841	1.8
26	321	43.3	105	.2	946	2.1
27	323	45.0	81	.2	1027	2.2
28	324	46.7	106	.2	1133	2.5
29	326	48.3	111	.2	1244	2.7
30	328	50.0	109	.2	1353	2.9
31	329	51.7	144	.3	1497	3.2
32	331	53.3	168	.4	1665	3.6
33	332	55.0	184	.4	1849	4.0
34	334	56.7	178	.4	2027	4.4
35	336	58.3	204	.4	2231	4.8
36	337	60.0	229	.5	2460	5.3
37	339	61.7	253	.5	2713	5.9
38	341	63.3	311	.7	3024	6.6
39	342	65.0	358	.8	3382	7.3
40	344	66.7	418	.9	3800	8.2
41	346	68.3	459	1.0	4259	9.2
42	348	70.0	512	1.1	4771	10.3
43	350	71.7	649	1.4	5420	11.8
44	352	73.3	762	1.7	6182	13.4
45	354	75.0	916	2.0	7098	15.4
46	356	76.7	1089	2.4	8187	17.8
47	358	78.3	1234	2.7	9421	20.4
48	360	80.0	1506	3.3	10927	23.7
49	363	81.7	1746	3.8	12673	27.5
50	366	83.3	2005	4.3	14678	31.8
51	368	85.0	2480	5.4	17158	37.2
52	372	86.7	2783	6.0	19941	43.2
53	375	88.3	3231	7.0	23172	50.3
54	379	90.0	3522	7.6	26694	57.9
55	383	91.7	4041	8.8	30735	66.7
56	388	93.3	4216	9.1	34951	75.8
57	395	95.0	4139	9.0	39090	84.8
58	404	96.7	3571	7.7	42661	92.5
59	419	98.3	2454	5.3	45115	97.8
60	438	100.0	993	2.2	46108	100.0

References

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting item bias. Baltimore: The Johns Hopkins University Press.

Chafin, A. E. (1983). Setting a standard for standard setting. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. NY: CBS College Publishing.

Gronlund, N. E. (1982). Constructing achievement tests (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Maryland State Department of Education (1980). Declared competencies index. Baltimore, MD: Author.

Maryland State Department of Education. (1981). Project basic instructional guide. Volume VI: Functional reading. Baltimore, MD: Author.

Maryland State Department of Education. (1982). Test specifications: Maryland functional reading tests. Baltimore, MD: Author.

Osterlind, S. J. (1985) Test item bias. Beverly Hills: Sage Publications.

Subkoviak, M. (1984) Estimating the reliability of mastery-nonmastery classifications. In R. A. Berk (Ed.), A guide to criterion-referenced test construction. Baltimore: The Johns Hopkins University Press.

Rudner, L. M. (1977). An evaluation of select approaches for biased item modification. Unpublished doctoral dissertation, Catholic University of America, Washington, D.C.

LWB28001